Bioinformatic analysis of host cell gene expression and chromatin accessibility in response to *Chlamydia trachomatis* infection

Regan J. Hayward

University of Technology Sydney

Faculty of Science,

School of Life Sciences, The i3 Institute

Supervised by:

Associate Professor Garry Myers

Associate Professor Wilhelmina Huston

A thesis submitted in fulfilment of the requirements for the degree:
Doctor of Philosophy

May 2020

Certificate of original authorship

I, Regan Hayward, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, School of Life Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 10th May 2020

Acknowledgements

The decision to return to university was based on a desire to challenge myself, but also to have fun and enjoy the experience along the way. This journey has spanned almost 8 years, and is now nearly complete. There have been plenty of high and lows over that time, but overall, I'm glad I made this decision and that I've been able to follow it through.

I've been fortunate to have the support from my amazing wife Giselle, who has listened, given advice, read through countless drafts and probably never wants to hear the words PhD or *Chlamydia* again! You have been amazing throughout this journey – thank you.

I'd also like to thank my family, who have supported in a range of different ways, particularly during these final months when they have had to travel internationally to allow me the time to focus on writing, especially since time has become a limited commodity with full time work.

I would like to thank my principal supervisor Garry Myers for the opportunity to work on this project and providing continuous support, guidance and mentorship over the course of my candidature. I'd also like to thank my co-supervisor Willa Huston, for her support and being available for many formal and informal chats. I would also like to thank all group members past and present for their support.

I would also like to say thank you to William King, Mark Thomas and Rami Mazraani who have been excellent colleagues and friends, who have made the course of my candidature all the more enjoyable.

Lastly, I'd like to thank the University of Technology Sydney and the i3 Institute for supporting me financially with a Doctoral Scholarship and the ability to travel to conferences locally and internationally.

Publications arising

Chapter 4 (Published)

Regan J. Hayward, James W. Marsh, Michael S. Humphrys, Wilhelmina M. Huston, Garry S. A. Myers. 2019: Early transcriptional landscapes of *Chlamydia trachomatis*-infected epithelial cells at single cell resolution: *Front. Cell. Infect. Microbiol*; vol 9: doi=10.3389/fcimb.2019.00392
Preprint: https://doi.org/10.1101/724641

Chapter 3 (Under peer-review and available as a preprint)

Regan J. Hayward, James W. Marsh, Michael S. Humphrys, Wilhelmina M. Huston, Garry S. A. Myers. 2019: Chromatin accessibility dynamics of *Chlamydia*-infected epithelial cells *bioRxiv*: https://doi.org/10.1101/681999
Preprint: https://doi.org/10.1101/681999

James W. Marsh, Regan Hayward, Amol Shetty, Anup Mahurkar, Michael S. Humphrys and Garry S. A. Myers. 2019. 'Dual RNA-Seq of *Chlamydia* and host cells', in Amanda Claire Brown (Ed.) *Chlamydia trachomatis*: Methods and Protocols. Humana Press. DOI: 10.1007/978-1-4939-9694-0

James W. Marsh, Regan J. Hayward, Amol C. Shetty, Anup Mahurkar, Michael S. Humphrys, Garry S. A. Myers; 2017, Bioinformatic analysis of bacteria and host cell dual RNA-sequencing experiments, *Briefings in Bioinformatics*, https://doi.org/10.1093/bib/bbx043
Preprint: https://doi.org/10.1101/098715

Table of contents

List of tables

List of figures

Abbreviations

3’UTR	3-Prime untranslated region
AB	Aberrant body
aRB	Aberrant reticulate body
ATAC-seq	Assaying for transposase-accessible chromatin sequencing
ATCC	American type culture collection
AutoML	Automated machine learning
BAM	Binary alignment map
CDC	Centre for disease control and prevention
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP–seq	Chromatin immunoprecipitation followed by sequencing
COREs	Clusters of open regulatory elements
CpG sites	Cytosine and guanine appearing consecutively on the same strand
CRISPR	Clustered regularly interspaced short palindromic repeats
DE	Differentially expressed
DMEM	Dulbecco's modified eagle medium
DMR	Differentially methylated regions
EB	Elementary body
EGFR	Epidermal growth factor receptor
EMT	Epithelial-mesenchymal transition
ENCODE	Encyclopaedia of DNA elements
FACS	fluorescence-activated cell sorting
FAIRE-seq	Formaldehyde-assisted isolation of regulatory elements
FBS	Fetal bovine serum

FDR	False discovery rate
FRAEM	Fluorescence-reported allelic exchange mutagenesis
FRiP	Fraction of reads in peaks
GEO	Gene expression omnibus
GO	Gene ontology
HDACs	Histone deacetylases
HEp-2	Human epithelial type 2 cells
Hi-C	Method to examine chromatin interactions from a 3D landscape
HPI	Hours post infection
HtrA	High temperature requirement protein A
HVG	Highly variable genes
IF	Intermediate filament
IFC	Integrated fluidic circuit
IGS	Institute for genome sciences
IGV	Integrative genomics viewer
KEGG	Kyoto encyclopaedia of genes and genomes
KLF	Krüppel-like-factors
KNN	K-nearest neighbour
LGV	Lymphogranuloma venereum
lincRNA	Long intergenic non-coding RNA
miRNA	MicroRNA
miscRNA	Miscellaneous RNA
MNase-seq	Micrococcal nuclease sequencing
MOMP	Major outer membrane protein
MOI	Multiplicity of infection

MSM	Men who have sex with men
MT	Metallothionein
MT rRNA	Mitochondrial rRNA
MTOC	Microtubule-organizing centre
NAAT	Nucleic acid amplification test
NCBI	National Centre for Biotechnology Information
ncRNA	Non-coding RNA
NF-*κB*	Nuclear factor-κB
NGS	Next generation sequencing
NUE	Nuclear effector
PA	Phosphatidic acid
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PID	Pelvic inflammatory disease
PolyA	Polyadenylated
QC	Quality control
RB	Reticulate body
RCA	Rolling circle amplification
ROS	Reactive oxygen species
Ribo-SPIA	RNA-based single-primer isothermal amplification
RLE	Relative log expression
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript, per million mapped reads
rRNA	Ribosomal RNA
SC3	Single-cell consensus clustering

scBS-seq	Single cell bisulfite sequencing
scRNA-seq	Single cell RNA sequencing
snoRNA	Small nucleolar RNA
SPG	Succinic phosphate glycine buffer
sRNA	Small non-coding RNA
S.D.	Standard deviation
STI	Sexually transmitted infection
T3SS	Type III secretion system
TF	Transcription factor
TFI	Tubal factor infertility
TMM	Trimmed mean of M-values
TNF	Tumour necrosis factor
TPM	Transcripts per kilobase million
tRNA	Transfer RNA
TSS	Transcription start site
TU	Transcription unit
TTS	Transcription termination site
UMI	Unique molecular identifiers
UV	Ultraviolet
VIM	Vimentin
WHO	World health organisation

Abstract

Chlamydia are Gram-negative, obligate intracellular bacterial pathogens responsible for a wide range of human and animal diseases. In humans, *Chlamydia trachomatis* is the most prevalent bacterial sexually transmitted infection (STI) worldwide and is the leading cause of trachoma (infectious blindness) in disadvantaged populations. If left untreated, infections can lead to more complex disease outcomes including infertility, ectopic pregnancy, epididymitis, prostatitis, and pelvic inflammatory disease. Due to widespread rates of infection and disease around the world and the associated economic costs, chlamydial infections remain a serious public health concern. All chlamydial species are defined by their unique intracellular developmental cycle. However, this has been a significant barrier restricting traditional molecular microbial investigation, such as transformation. As a result, we still do not have a comprehensive understanding of chlamydial gene function, particularly secreted effector proteins that modulate many host cell interactions. In the absence of a reliable and efficient transformation system, next generation sequencing (NGS) approaches enable the recovery of genome-wide expression patterns from a chlamydial or host point of view to aid in uncovering these functions and interactions.

To help with further characterisation and identification of these host-chlamydial interactions, this work applied three novel NGS approaches using *in vitro* models of infection with *C. trachomatis*. Chapter 3 examines chromatin accessibility dynamics across the developmental cycle (1, 12, 24 and 48 hours) to identify epigenomic changes to host cells; Chapter 4 utilises single cell RNA-sequencing (scRNA-seq) from host cells to examine early developmental time points (3, 6 and 12 hours); and Chapter 5 simultaneously examines host and chlamydial expression (dual RNA-seq) from two time points (1 and 24 hours), with an experimental design aimed to examine different depletion techniques and to optimise the ratio of EBs per cell for infection models.

Examination of the host cell epigenome identified both conserved and distinct temporal changes genome-wide. Differentially accessible chromatin regions were associated with immune responses, re-direction of host cell nutrients, intracellular signalling, cell-cell adhesion, extracellular matrix, metabolism and apoptosis. Temporally enriched transcription factors identified a novel family of Krüppel-like-factors (KLFs) which are ubiquitously expressed in reproductive tissues and associated with a variety of uterine pathologies.

Analyses from scRNA-seq highlight infection-specific host cell biology, including two distinct clusters separating 3 hour cells from 6 and 12 hours. Pseudotime analysis identified a possible infection-specific cellular trajectory for *Chlamydia*-infected cells, and differential expression identified temporally expressed genes involved with cell cycle regulation, innate immune responses, cytoskeletal components, lipid biosynthesis and cellular stress.

Dual RNA-seq analysis showed that combining depletion methods (polyA and rRNA) increases the capture rate of chlamydial transcripts, but negatively impacts host-cell expression. Different MOIs (0.1, 1 and 10) highlighted that an MOI of 10 captures significantly more transcripts and is more beneficial for capturing chlamydial transcripts.

Overall, this work highlights the complex nature of chlamydial infections, uncovering novel biological functions and regulatory activities. These results and analyses also provide further considerations and improvements for future *in vitro* experiments, but also enable the application of these genome-scale techniques to the investigation of complex disease models *in vivo* and in human tissues *ex vivo*.